

Joint Entropy Minimization for Learning in Nonparametric Framework

Parvez Ahammad



Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2006-85

<http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-85.html>

June 9, 2006

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 09 JUN 2006		2. REPORT TYPE		3. DATES COVERED 00-00-2006 to 00-00-2006	
4. TITLE AND SUBTITLE Joint Entropy Minimization for Learning in Nonparametric Framework				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of California at Berkeley,Electrical Engineering and Computer Sciences,Berkeley,CA,94720				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 41	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Copyright © 2006, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Joint Entropy Minimization for Learning in Nonparametric Framework

by Parvez Ahammad

parvez@eecs.berkeley.edu

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences,
University of California at Berkeley,
in partial satisfaction of the requirements for the degree of

Master of Science, Plan II.

Approval for the Report and Comprehensive Examination:

Committee:

Prof. Shankar S. Sastry
Research Advisor

Date

* * * * *

Prof. Jitendra Malik
Second Reader

Date

Contents

1	Overview	1
2	Joint MRI Bias Removal	3
2.1	Introduction	3
2.2	The Image Model and Problem Formulation	5
2.3	Joint Bias Removal Algorithm	8
2.4	Experiments	10
2.5	Summary and Future Work	11
3	Joint Alignment of Drosophila Imaginal Discs	15
3.1	Introduction	15
3.1.1	Motivation and Problem Definition	15
3.1.2	Previous Work	17
3.2	Proposed Approach	18
3.2.1	Image Model for Drosophila Imaginal Discs	18
3.3	Extracting Tissue Shapes via Segmentation	19
3.4	Joint Nonparametric Shape Learning and Alignment	22
3.4.1	Shape Learning	23
3.4.2	Joint Alignment	24
3.5	Stain Scoring	29
3.6	Summary and Future Work	29

To **God**

Acknowledgements

Many people have helped me along the way, and I owe my deep gratitude to them. It is not practical to mention each of them here, so I will mention the few who helped me during the phase when I worked on the research discussed in this report.

First, I would like to thank my academic advisor, Shankar Sastry, for having faith in my ability (even when I was fairly new to the field of computer vision), for his generous support, and for providing unmatched intellectual freedom to try new things out. His erudition, mathematical expertise, visionary insight into new directions, and dedication to standards of academic excellence are exemplary. I have learned a lot from him so far.

I have immensely enjoyed learning from Jitendra Malik, both through the class he taught, and his insightful comments during vision group meetings. His knowledge in the field of computer vision is simply remarkable and I have benefited from it. I thank him for being the second reader for this report.

Erik Learned-Miller is a core reason for the existence of this work. He introduced me to these ideas, helped me when I faltered, and patiently guided me to the stage where I could apply the ideas into new domains on my own. It has been a pleasure knowing him, both as a person and as a researcher. Cyrus Harmon has also been a wonderful collaborator during this period. It has been a lot of fun working with him. These two have co-authored papers with me, which form the basis for this report.

I would like to thank Sue Celniker, Richard Weiszmann and the Berkeley Drosophila Genome Project for help with data generation for the Drosophila imaginal discs, and Dr. Terrie Inder and Dr. Simon Warfield for providing the infant brain images for this work. The infant brain images were obtained under NIH grant P41 RR13218. This research was partially supported by Army Research Office grant DAAD 19-02-1-0383.

Although they didn't directly contribute to this work, some people have been instrumental in improving my understanding of various concepts, and in keeping my life sane when things weren't exactly going right. On this note, I want to thank (names listed in alphabetical order): Alessandro Abate, Aaron Ames, Ruzena Bajcsy, Minghua Chen, Phoebus Chen, Sharon Core, Mikael Eklund, Christopher Geyer, Richard Groff, Ann Hammonds, Rajasekhar and Padma Kamada, Ali Lakhia, Chung-Shik Lee, Vivian Leung, Carmel Majidi, Marci Meingast, Songhwai Oh, Walter Richter, Tanya Roosta, Gerald Rubin, Jonathan Sprinkle, Mary Margaret Sprinkle and Bin Yu. I express my sincere gratitude to my family and close friends for their understanding and support through these years.

List of Figures

2.1	Illustration of the effect of bias in MR images	4
2.2	Pixel stacks in MR images	6
2.3	Entropy of the pixel stack distribution	7
2.4	Sine/Cosine Basis Fields	8
2.5	BrainWeb Sample Images	10
2.6	Experimental results for BrainWeb Images with different but known bias fields . . .	13
2.7	Results for Infant Brain MR image set	14
3.1	A typical set of in situ stained Drosophila imaginal disc images	16
3.2	Data flow in our proposed approach.	17
3.3	Example segmentation results using the combined segmentation procedure.	19
3.4	Mean shape images from the learning stage.	21
3.5	Alignment results for wing discs.	25
3.6	Alignment results for haltere discs.	26
3.7	Alignment results for leg discs.	27
3.8	Alignment results for eye discs.	28
3.9	Stain patterns in Drosophila imaginal disc images: unaligned vs. aligned.	30

Chapter 1

Overview

In computer vision and image processing fields, we come across problems where we have multiple examples of data that are corrupted, but all generated from some unknown underlying model. An interesting challenge here is to estimate the underlying generative model without making any prior assumptions on the model, while simultaneously estimating the various noises or transformations affecting this model, thus resulting in the noisy or transformed data as observed in real life scenarios.

We investigate an information theoretic formulation to address these issues [19], especially in two practical biomedical scenarios: the first is to estimate and rectify the random field (RF) bias in magnetic resonance (MR) images [16], and the second is to align noisy images of stained *Drosophila* specimens to facilitate accurate quantitative analysis of gene expression [2]. The first scenario demonstrates the applicability of this framework to non-spatial transformations, whereas the second scenario demonstrates the applicability of the framework to shape transformations in affine space. This framework is easily extendable to problems where both spatial and non-spatial parameters need to be learned.

The correction of RF bias in MR images is an important problem in medical image processing [12, 26]. Most previous approaches have used a maximum likelihood method to increase the likelihood of the pixels in a *single* image by making small changes to a correcting bias field. The likelihood is defined either in terms of a pre-existing tissue model, or non-parametrically in terms of the image's own pixel values. In this work, we suggest a new approach [16] in which we *simultaneously* rectify the bias from a *set of images* of the same anatomy, but from different patients. We use the statistics from the same location across different images, rather than within an image, to eliminate all but a single constant bias field from all of the images. We then use a second procedure to remove this constant bias field simultaneously from all of the images. Our method uses a source of structure, namely *statistics across images*, that goes completely unused in other techniques. Our formulation cleanly extends to application of spatial transformations as well as non-spatial transformations based on the bias field application according to the image formation model.

To compare spatial patterns of gene expression, one must analyze a large number of images as current methods are only able to measure expression patterns of a small number of genes at a time. Bringing images of corresponding tissues into alignment is a critical first step in making a meaningful comparative analysis of these spatial patterns. Significant image noise (and clutter)

and variability in the shapes make it hard to pick a canonical shape model. In our work [2], we address these problems by combining segmentation and unsupervised shape learning algorithms. We first segment images to acquire structures of interest, then jointly align the shapes of these acquired structures using an unsupervised nonparametric maximum likelihood algorithm along the lines of ‘congealing’ [17], while simultaneously learning underlying shape model and the associated transformations. The learned transformations are applied to corresponding gene expression images to bring them into alignment in one step. We demonstrate the results for images of various classes of *Drosophila* imaginal discs and discuss the methodology used for making quantitative analysis of these spatial gene expression patterns.

Contributions: This work makes two contributions. First contribution is to extend the entropy minimization framework [19] to gray scale MRI images for modeling and removing random field bias in MR images. Second contribution is to combine segmentation and ‘congealing’ to solve alignment problem in spatial gene expression data analysis for *Drosophila* imaginal discs.

Chapter 2

Joint MRI Bias Removal

2.1 Introduction

Magnetic Resonance (MR) imaging is a powerful noninvasive imaging modality that has experienced rapid growth over the past decade. Standard applications of MR include diagnostic imaging studies of the central nervous system and musculo-skeletal system [20]. There are a number of artifacts that can arise in the MR imaging process and make subsequent analysis very challenging. Possibly the most drastic visual effect is the intensity inhomogeneity caused by the spatially varying signal response of the electrical coil that receives the MR signal. This coil inhomogeneity results in a multiplicative gain field that biases the observed signal from the true underlying signal [12]. This problem is illustrated in Figure 2.1. When a patient is imaged in the MR scanner, the goal is to obtain an image which is a function solely of the underlying tissue (left of Figure 2.1). However, typically the desired anatomical image is corrupted by a multiplicative bias field (2nd image of Figure 2.1) that is caused by engineering issues such as imperfections in the radio frequency coils used to record the MR signal. The result is a corrupted image (3rd image of Figure 2.1). Most of the diagnostic MR image processing procedures operate on the intensity values obtained in MR images. MR images are constructed from electromagnetic responses and are captured using coils of wire. These intensities are corrupted both by random noise as well as systematic electromagnetic effects. The latter are collectively known as bias fields or intensity inhomogeneities. The bias in this case is a multiplicative bias rather than an additive bias which is more common. The term bias is used because the intensity inhomogeneity is a systematic effect and not a random effect. Image processing in general relies on the intensity values and can be significantly impaired by imperfections in the image collection process. Both the noise and the bias can confuse automated image processing algorithms, and it is highly desirable to minimize both as much as possible. The goal of MR bias correction is to estimate the uncorrupted image from the corrupted image. The bias correction problem is currently a challenging one and is very widely studied. The importance of this problem will increase as MR magnets increase in strength and electromagnetic effects become more and more pronounced.

A variety of statistical methods have been proposed to address this problem. Wells et al. [26]

¹This chapter is a more detailed version of [16]

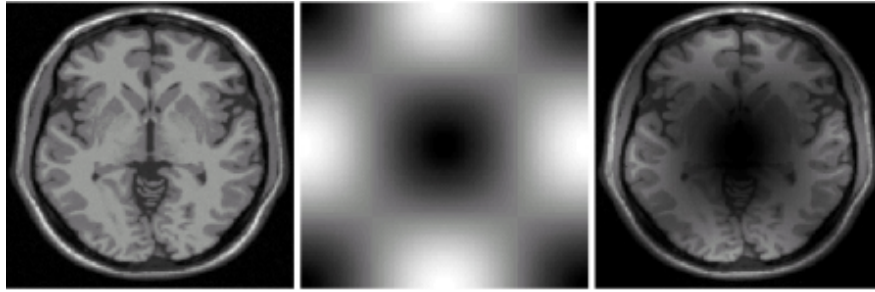


Figure 2.1: Illustration of the effect of bias in MR images

On the left is a mid-axial MRI scan of the human brain with little or no bias field. In the center is a simulated low-frequency bias field. It has been exaggerated for ease of viewing. On the right is the result of pixelwise multiplication of the image by the bias field. The goal of MR bias correction is to recover the low-bias image on the left from the biased image on the right.

developed a statistical model using a discrete set of tissues, with the brightness distribution for each tissue type (in a bias-free image) represented by a one-dimensional Gaussian distribution. An expectation-maximization (EM) procedure was then used to simultaneously estimate the bias field, the tissue type, and the residual noise. While this method works well in many cases, it has several drawbacks:

1. Models must be developed a priori for each type of acquisition (for each different setting of the MR scanner), for each new area of the body, and for different patient populations (like infants and adults).
2. Models must be developed from *bias-free* images, which may be difficult or impossible to obtain in many cases.
3. The model assumes a fixed number of tissues, which may be inaccurate. For example, during development of the human brain, there is continuous variability between gray matter and white matter.

In addition, a discrete tissue model does not handle so-called partial volume effects in which a pixel represents a combination of several tissue types. This occurs frequently since many pixels occur at tissue boundaries. Non-parametric approaches have also been suggested, as for example by Viola [25]. In that work, a non-parametric model of the tissue was developed from a single image. Using the observation that the entropy of the pixel brightness distribution for a single image is likely to increase when a bias field is added, Viola's method postulates a bias-correction field by minimizing the entropy of the resulting pixel brightness distribution. This approach addresses several of the problems of fixed-tissue parametric models, but has its own drawbacks:

1. The statistical model may be weak, since it is based on data from only a single image.
2. There is no mechanism for distinguishing between certain low-frequency image components and a bias field. That is, the method may mistake signal for noise in certain cases when removal of the true signal reduces the entropy of the brightness distribution.

We shall show that this is a problem in real medical images.

The method we present overcomes or improves upon problems associated with both of these methods and their many variations (see, e.g., [12] for recent techniques). It models tissue brightness non-parametrically, but uses data from multiple images to provide improved distribution estimates and alleviate the need for bias-free images for making a model. It is also conditional on spatial location, taking advantage of a rich information source ignored in other methods. Experimental results demonstrate the effectiveness of our method.

2.2 The Image Model and Problem Formulation

We assume we are given a set I of observed images I_i with $1 \leq i \leq N$, as shown on the left side of Figure 2.2. Each of these images is assumed to be the product of some bias-free image L_i and a smooth bias field $B_i \in \Phi$. We shall refer to the bias-free images as latent images (also called *intrinsic images* by some authors). The set of all latent images shall be denoted L and the set of unknown bias fields B . Then each observed image can be written as the product $I_i(x, y) = L_i(x, y) * B_i(x, y)$, where (x, y) gives the pixel coordinates of each point, with P pixels per image. Consider again Figure 2.2. A *pixel-stack* through each image set is shown as the set of pixels corresponding to a particular location in each image (not necessarily the same tissue type). Our method relies on the principle that the pixel-stack values will have lower entropy when the bias fields have been removed. Figure 2.3 shows the simulated effect, on the distribution of values in a pixel-stack, of adding different bias fields to each image.

The latent image generation model assumes that each pixel is drawn from a fixed distribution $p_{x,y}(\cdot)$ which gives the probability of each gray value at the the location (x, y) in the image. Furthermore, we assume that all pixels in the latent image are independent, given the distributions from which they are drawn. It is also assumed that the bias fields for each image are chosen independently from some fixed distribution over bias fields. Unlike most models for this problem which rely on statistical regularities within an image, we take a completely orthogonal approach by assuming that pixel values are independent given their image locations, but that pixel-stacks in general have low entropy when bias fields are removed.

We formulate the problem as a maximum a posteriori (MAP) problem, searching for the most probable bias fields given the set of observed images. Letting Ψ represent the 25-dimensional product space of smooth bias fields (corresponding to the 25 basis images of Figure 2.1), we wish to find $\arg \max_{B \in \Psi} P(B|I)$. We define $\hat{\Theta}$ as:

$$\hat{\Theta} = \arg \max_{B \in \Phi} P(B|I). \quad (2.1)$$

Using Bayes' rule and ignoring the constant denominator, we can write it as:

$$\hat{\Theta} = \arg \max_{B \in \Phi} P(I|B)P(B). \quad (2.2)$$

$$(2.3)$$

Assuming uniform prior over the allowed bias fields, we can write this as:

$$\hat{\Theta} = \arg \max_{B \in \Phi} P(I|B). \quad (2.4)$$

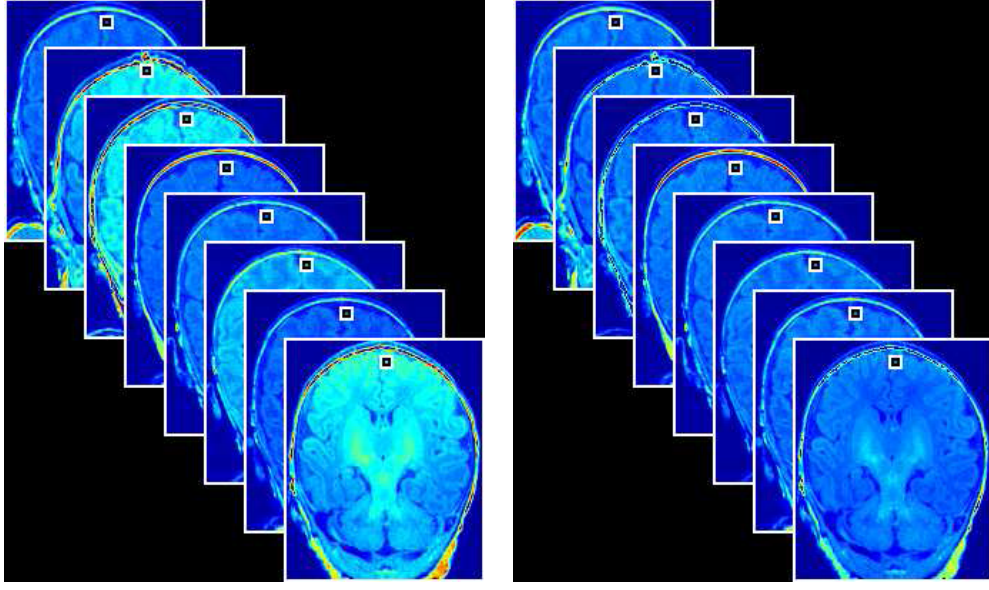


Figure 2.2: Pixel stacks in MR images

On the left are a set of mid-coronal brain images from eight different infants, showing clear signs of bias fields. A pixel-stack, a collection of pixels at the same point in each image, is represented by the small square near the top of each image. Although there are probably no more than two or three tissue types represented by the pixel-stack, the brightness distribution through the pixel-stack has high empirical entropy due to the presence of different bias fields in each image. **On the right** are a set of images that have been corrected using our bias field removal algorithm. While the images are still far from identical, the pixel-stack entropies have been reduced by mapping similar tissues to similar values in an *unsupervised fashion*, i.e. without knowing or estimating the tissue types.

Our method can be easily altered to incorporate non-uniform prior as well. The probability of an observed image given a particular bias field is the same as the probability of the latent image associated with that observed image and bias field. This can be expressed as

$$\hat{\Theta} = \arg \max_{B \in \Phi} P(L(I, B)) \quad (2.5)$$

$$= \arg \max_{B \in \Phi} \prod_{x,y} \prod_{i=1}^N p_{x,y}(L_i(x, y)) \quad (2.6)$$

$$= \arg \max_{B \in \Phi} \sum_{x,y} \sum_{i=1}^N p_{x,y}(L_i(x, y)) \quad (2.7)$$

$$(2.8)$$

At each pixel, the empirical mean of the log probability can be approximated with the negative entropy of the underlying distribution at that pixel. This can be written as:

$$\hat{\Theta} \approx \arg \min_{B \in \Phi} \sum_{x,y} H(p_{x,y}). \quad (2.9)$$

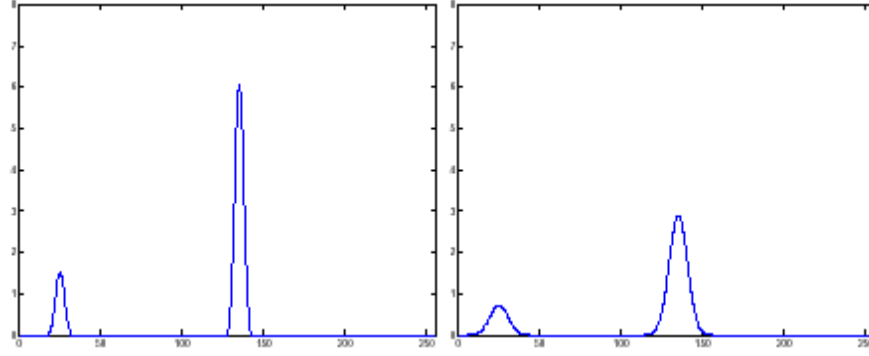


Figure 2.3: Entropy of the pixel stack distribution

On the left is a simulated distribution from a pixel-stack taken through a particular set of bias-free mid-axial MR images. The two sharp peaks in the brightness distribution represent two tissues which are commonly found at that particular pixel location. **On the right** is the result of adding an independent bias field to each image. In particular, the spread, or entropy, of the pixel distribution increases. In this work, we seek to remove bias fields by seeking to reduce the entropy of the pixel-stack distribution to its original state.

Here H is the Shannon entropy defined as $H = \mathbb{E}_p(-\log p_{x,y})$. We use the entropy estimator of Vasicek [24] to directly estimate this entropy from the samples in the pixel-stack, without ever estimating the distributions $p_{x,y}$ explicitly.

$$\hat{\Theta} \approx \arg \min_{B \in \Phi} \sum_{x,y} \hat{H}_{Vasicek}(L_1(x,y), \dots, L_N(x,y)) \quad (2.10)$$

$$\hat{\Theta} = \arg \min_{B \in \Phi} \sum_{x,y} \hat{H}_{Vasicek}\left(\frac{I_1(x,y)}{B_1(x,y)}, \dots, \frac{I_N(x,y)}{B_N(x,y)}\right). \quad (2.11)$$

Mathematically, this ML estimation can be formulated as solving an optimization problem. Noting the fact that the priors aren't actually uniform on different bias fields, we need to write a compensated objective function for our optimization Ψ defined as

$$\Psi \doteq \sum_{B \in \Phi} \sum_{x,y} \hat{H}_{Vasicek}(L_1(x,y), \dots, L_N(x,y)) + \sum_{i=1}^N |\gamma^i| \quad (2.12)$$

where γ^i are the vectors of bias field parameters (Equation 2.13), and $|\cdot|$ is some norm (or penalty term or regularization term) on these vectors. Ψ is called the penalized pixel-wise entropy [19]. The approximation in Equation 2.9 becomes an equality as N grows large by the law of large numbers, while the consistency of Vasicek's entropy estimator [6] implies that Equation 2.10 also goes to equality with large N . (Please refer to [6] for a review of entropy estimators).

2.3 Joint Bias Removal Algorithm

Using these ideas, it is straightforward to construct algorithms for joint bias field removal. As mentioned above, we chose to optimize Equation 2.11 over the set of band-limited bias fields. To do this, we parameterize the set of bias fields using the sine/cosine basis images shown on the right of Figure 2.4:

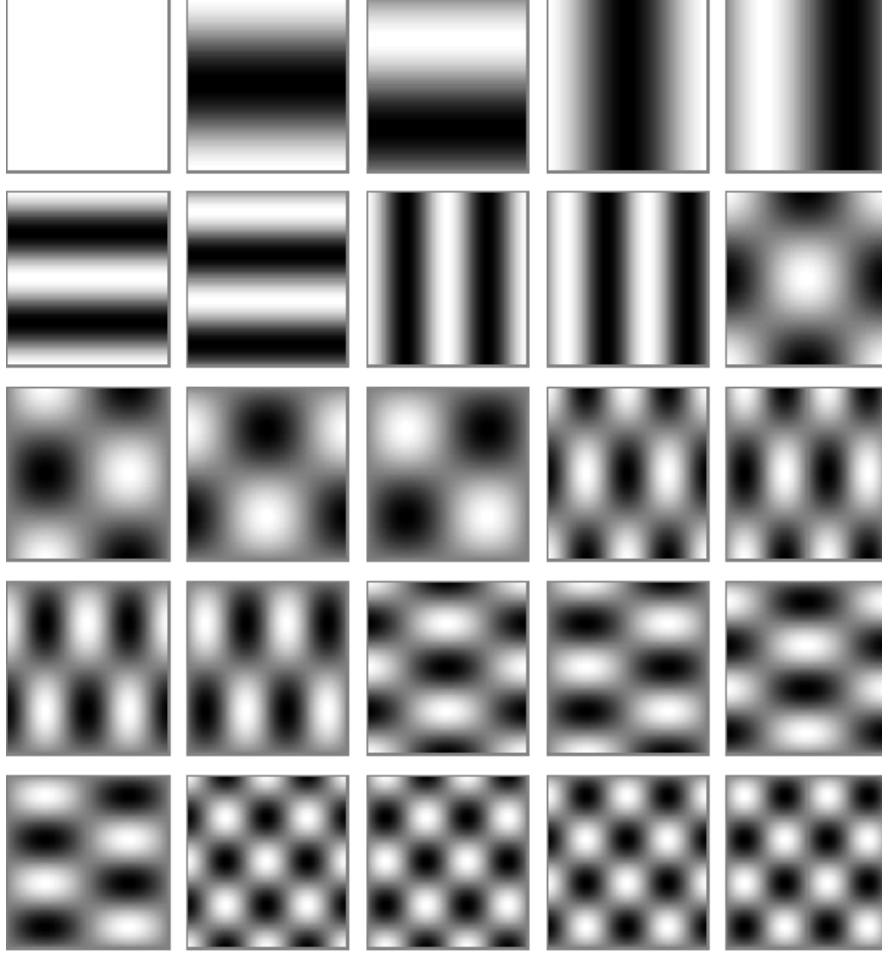


Figure 2.4: Sine/Cosine Basis Fields

25 sine/cosine basis fields are shown here, that are combined to construct band-limited bias fields using Equation 2.13

$$B_i = \sum_1^{25} \gamma_j^i \phi_j(x, y). \quad (2.13)$$

We optimize Equation 2.11 by simultaneously updating the bias field estimates (taking a step along the numerical gradient) for each image to reduce the overall entropy. That is, at time step t , the coefficients γ_j for each bias field are updated using the latent image estimates and entropy estimates

from time step $t - 1$. After all γ s have been updated, a new set of latent images and pixel-stack entropies are calculated, and another gradient step is taken. Though it is possible to do a full gradient descent to convergence by optimizing one image at a time, the optimization landscape tends to have more local minima for the last few images in the process. The appeal of our joint gradient descent method, on the other hand, is that the ensemble of images provides a natural smoothing of the optimization landscape in the joint process. It is in this sense that our method is *multi-resolution*, proceeding from a smooth optimization in the beginning to a sharper one near the end of the process.

We now summarize the algorithm:

1. Initialize the bias field coefficients for each image to 0, with the exception of the coefficient for the DC-offset (the constant bias field component), which is initialized to 1. Initialize the gradient descent step size δ to some value.
2. Compute the summed pixelwise entropies for the set of images with initial *neutral* bias field corrections. (See below for method of computation.)
3. Iterate the following loop until no further changes occur in the images.
 - (a) For each image:
 - i. Calculate the numerical gradient $\nabla_{\gamma_i} \Psi$ of equation 2.11 with respect to the bias field coefficients (γ_j s) for the current image.
 - ii. Set $\gamma = \gamma + \delta \nabla_{\gamma_i} \Psi$.
 - (b) Update δ (reduce its value according to some reasonable update rule such as the Armijo rule [8]).

Upon convergence, it is assumed that the entropy has been reduced as much as possible by changing the bias fields, unless one or more of the gradient descents is stuck in a local minimum. Empirically, the likelihood of sticking in local minima is dramatically reduced by increasing the number of images (N) in the optimization. In our experiments described below with only 21 real infant brains, the algorithm appears to have found a global minimum of all bias fields, at least to the extent that this can be discerned visually.

Note that for a set of *identical* images, the pixel-stack entropies are not increased by multiplying each image by the same bias field (since all images will still be the same). More generally, when images are approximately equivalent, their pixel-stack entropies are not significantly affected by a *common* bias field, i.e. one that occurs in all of the images.¹ This means that the algorithm cannot, in general, eliminate all bias fields from a set of images, but can only set all of the bias fields to be equivalent. We refer to any constant bias field remaining in all of the images after convergence as the *residual bias field*. Fortunately, there is an effect that tends to minimize the impact of the residual bias field in many test cases. In particular, the residual bias field tends to

¹Actually, multiplying each image by a bias field of small magnitude can artificially reduce the entropy of a pixel-stack, but this is only the result of the brightness values shrinking towards zero. Such artificial reductions in entropy can be avoided by normalizing a distribution to unit variance between iterations of computing its entropy, as is done in this work.

consist of components for each γ_j that approximate the mean of that component across images. For example, if half of the observed images have a positive value for a particular components coefficient, and half have a negative coefficient for that component, the residual bias field will tend to have a coefficient near zero for that component. Hence, the algorithm naturally eliminates bias field effects that are non-systematic, i.e. that are not shared across images.

If the same type of bias field component occurs in a majority of the images, then the algorithm will not remove it, as the component is indistinguishable, under our model, from the underlying anatomy. In such a case, one could resort to within-image methods to further reduce the entropy. However, there is a risk that such methods will remove components that actually represent smooth gradations in the anatomy. This can be seen in the bottom third of Figure 2.7), and will be discussed in more detail below.

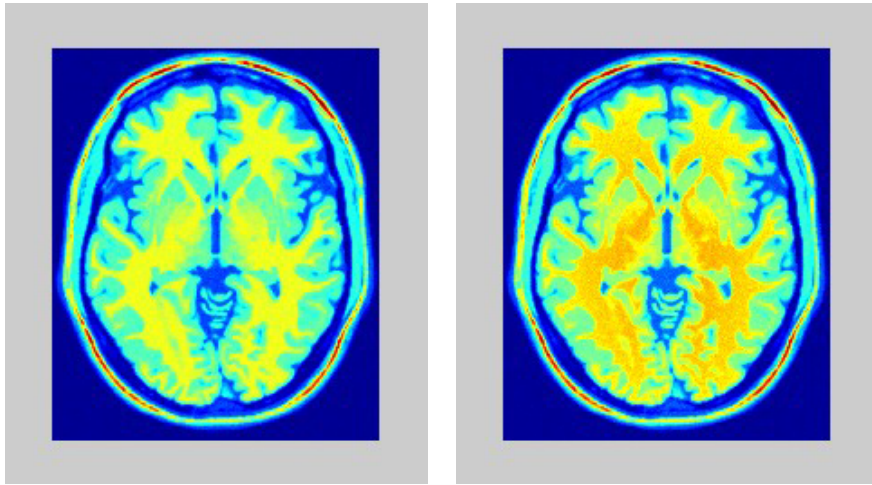


Figure 2.5: BrainWeb Sample Images

2.4 Experiments

To test our algorithm, we ran two sets of experiments, the first on synthetic images for validation, and the second on real brain images. We obtained synthetic brain images from the BrainWeb project [11, 1] such as the ones shown in Figure 2.5). The left image is a clean brain phantom with no bias, and the bottom image is a brain phantom corrupted by some bias field. These images can be considered *idealized* MR images in the sense that the brightness values for each tissue are constant (up to a small amount of manually added isotropic noise). That is, they contain no bias fields (the left image in Figure 2.5). The initial goal was to ensure that our algorithm could remove synthetically added bias fields, in which the bias field coefficients were known. Using K copies of a single *latent* image, we added known but different bias fields to each one. In our experiments, for as few as five images, we could reliably recover the known bias field coefficients, up to a fixed offset for each image, to within 1% of the power of the original bias coefficients. We show the results on BrainWeb synthetic images in Figure 2.6). As expected, when the bias removal is done, the images look more like each other, since the latent image among all the images is the same in this

experiment.

More interesting are the results on real images, in which the latent images come from different patients. We obtained 21 pre-registered ² infant brain images (top of Figure 2.7) from Brigham and Womens Hospital in Boston, Massachusetts. Large bias fields can be seen in many of the images. Probably the most striking is a *ramp-like* bias field in the sixth image of the second row. (The top of the brain is too bright, while the bottom is too dark.) Because the brain’s white matter is not fully developed in these infant scans, it is difficult to categorize tissues into a fixed number of classes as is typically done for adult brain images; hence, these images are not amenable to methods based on specific tissue models developed for adults (e.g. [26]).

The middle third of Figure 2.7 shows the results of our algorithm on the infant brain images. (These results must be viewed in color on a good monitor to fully appreciate the results.) While a trained technician can see small imperfections in these images, the results are remarkably good. All major bias artifacts have been removed.

It is interesting to compare these results to a method that reduces the entropy of each image individually, without using constraints between images. Using the results of our algorithm as a starting point, we continued to reduce the entropy of the pixels *within* each image (using a method akin to Violas [25]), rather than across images. These results are shown in the bottom third of Figure 2.7. Carefully comparing the central brain regions in the middle section of the figure and the bottom section of the figure, one can see that the butterfly shaped region in the middle of the brain, which represents developing white matter, has been suppressed in the lower images. This is most likely because the entropy of the pixels *within a particular image* can be reduced by increasing the bias field *correction* in the central part of the image. In other words, the algorithm strives to make the image more uniform by removing the bright part in the middle of the image. However, our algorithm, which compares pixels across images, does not suppress these real structures, since they occur across images. Hence coupling across images can produce superior results.

2.5 Summary and Future Work

The idea of minimizing pixelwise entropies to remove nuisance variables from a set of images is not new. In particular, Miller et al. [17, 19] presented an approach they call *congealing* in which the sum of pixelwise entropies is minimized by separate affine transforms applied to each image. Our method can thus be considered an extension of the congealing process to non-spatial transformations. Combining such approaches to do registration and bias removal simultaneously, or registration and lighting rectification of faces, for example, is an obvious direction for future work.

This work uses information unused in other methods, i.e. *information across images*. This suggests an iterative scheme in which both types of information, both within and across images, are used. Local models could be based on weighted neighborhoods of pixels, *pixel cylinders*, rather than single

²It is interesting to note that registration is not strictly necessary for this algorithm to work. The proposed MAP method works under very broad conditions, the main condition being that the bias fields do not span the same space as parts of the actual medical images. It is true, however, that as the latent images become less registered or differ in other ways, that a much larger number of images is needed to get good estimates of the pixel-stack distributions.

pixel-stacks, in sparse data scenarios. For *easy* bias correction problems, such an approach may be overkill, but for difficult problems in bias correction, where the bias field is difficult to separate from the underlying tissue, as discussed in [12], such an approach could produce critical extra leverage.

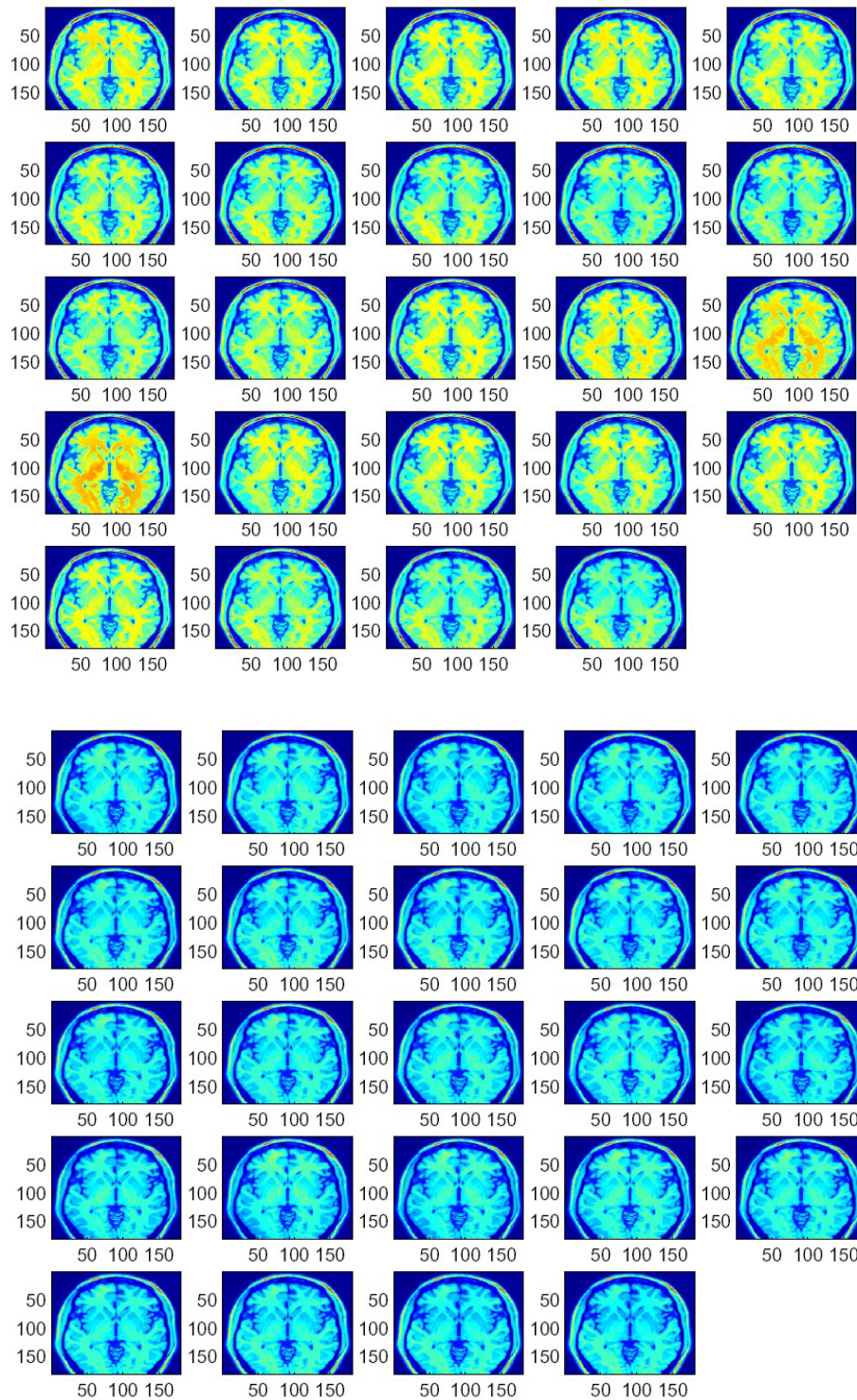


Figure 2.6: Experimental results for BrainWeb Images with different but known bias fields
Left: Brainweb images before bias removal. **Right:** Brainweb images after bias removal.
NOTE: This image must be viewed in color (preferably on a bright display) for full effect.

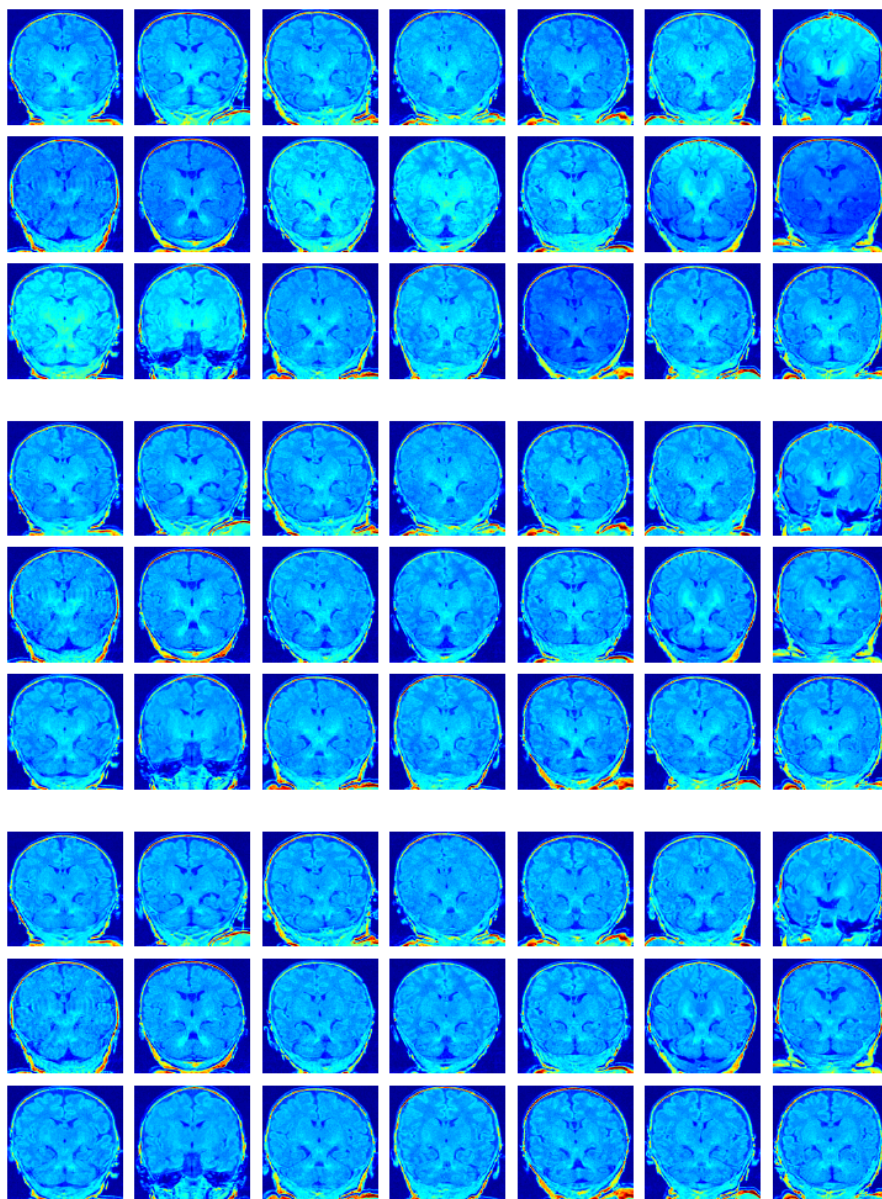


Figure 2.7: Results for Infant Brain MR image set

NOTE: This image must be viewed in color (preferably on a bright display) for full effect. **Top.** Original infant brain images. **Middle.** The same images after bias removal with our algorithm. Note that developing white matter (butterfly-like structures in middle brain) is well-preserved. **Bottom.** Bias removal using a single image based algorithm. Notice that white matter structures are repressed.

Chapter 3

Joint Alignment of *Drosophila* Imaginal Discs

3.1 Introduction

Microarray technologies have enabled researchers to measure levels of expression of large numbers of individual genes in a single experiment, thereby providing a greatly enhanced view of the genes that are active at a given time in specific tissues [18, 3]. These techniques generally require a large tissue sample and, on their own, provide no information about the spatial patterns of gene expression. In situ hybridization of tissues with labeled probes for individual genes facilitates the measurement of precise spatial patterns of gene expression at high resolution. However, in situ hybridization can only be used to measure a limited number of genes at a time, usually one. It is desirable to be able to measure the spatial expression patterns of large numbers of genes and to be able to compare, cluster, and classify patterns of expression. Current experimental strategy entails setting up a high-throughput production system for the generation of large numbers of images which can then be processed by either human or computer. Many dipteran organisms, including the fruit fly *Drosophila melanogaster*, have a three-stage life-cycle in which the insect begins as an embryo, becomes a larva and metamorphoses into an adult, also known as an imago. The primordial tissues that will become the exoskeleton of the adult insect, or imago, called imaginal discs, are segregated from the larval tissues in embryogenesis and are maintained in sac-like structures that have a morphology similar to that of a flattened balloon[5].

3.1.1 Motivation and Problem Definition

The tissue structures in typical *Drosophila* imaginal discs have significant intra and inter-class variability in size, shape and stain patterns. Thus, shape models learned from one class cannot be used for another tissue class without making significant changes in the processing pipeline if one were to use model based alignment algorithms. Furthermore, there are far fewer identified

¹This chapter is a more detailed version of [2]

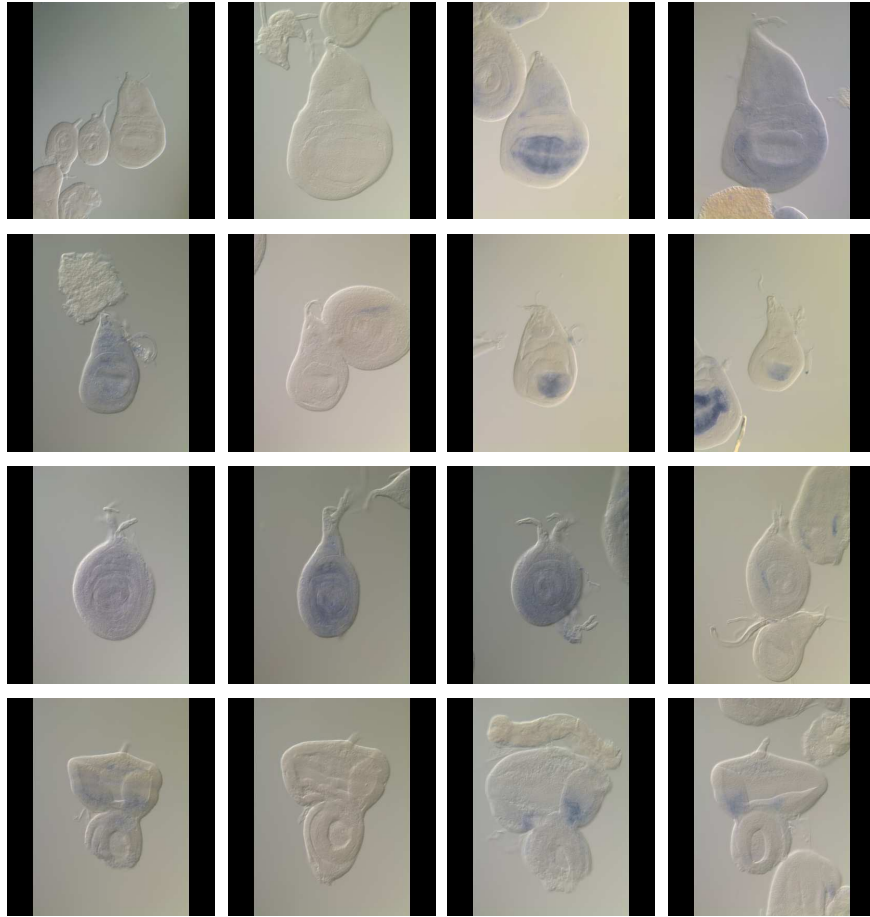


Figure 3.1: A typical set of in situ stained *Drosophila* imaginal disc images

Each row shows a different tissue class. First row: wing discs, second row: haltere discs, third row: leg discs, fourth row: eye discs. Notice the significant inter-class and intra-class variability both in the shapes, sizes and the stain patterns. Still, expert human operators have an implicit idea of what a canonical model of this shape must look like for each tissue class.

and named morphological parts or regions in *Drosophila* imaginal discs than in the developing *Drosophila* embryo, for example. It is difficult to make any parametric or model based assumptions for tissue shapes given these limitations. Manual annotation and curation are extremely time-consuming and costly in high-throughput spatial gene expression analysis experiments. It is highly desirable to have a processing pipeline that can operate for various imaginal disc tissues such as wings, halteres, legs, eyes, etc., (shown in Figure 3.1) without significant re-structuring. In this work, we propose a simple yet effective computational methodology (Section 3.2) that addresses these demands of high-throughput systems for the analysis of spatial gene expression by combining segmentation and nonparametric alignment algorithms.

Given an input ensemble of noisy *Drosophila* imaginal disc images of a given tissue class, our goal is to learn the underlying shape model of the tissue nonparametrically while bringing the

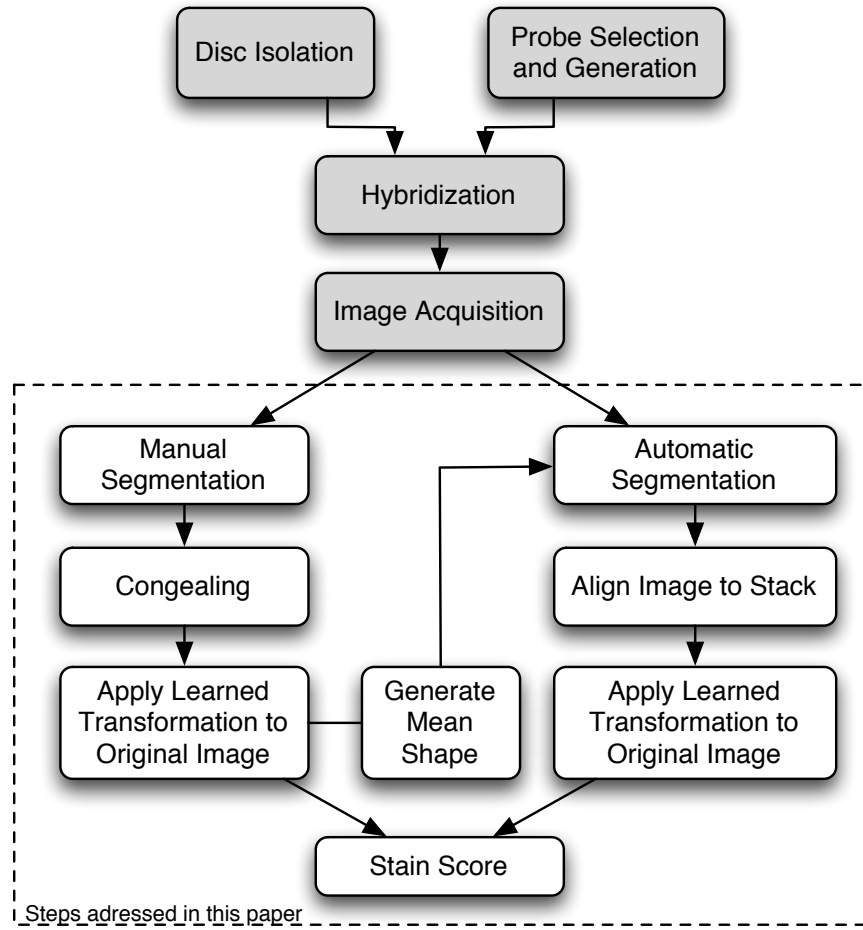


Figure 3.2: Data flow in our proposed approach.

given images into alignment. This alignment greatly facilitates quantitative stain scoring analysis (Section 3.5) on the imaginal disc images. We demonstrate the segmentation (Section 3.3) and alignment (Section 3.4) results for various tissue classes of *Drosophila* imaginal discs and discuss the results (Section 3.6).

3.1.2 Previous Work

Large-scale studies of patterns of gene expression in *Drosophila* have been performed using DNA microarrays both on whole organisms [4] and individual tissues such as imaginal discs. Klebes et al. compared differential gene expression in different imaginal discs and between imaginal discs and non-disc tissue [14]. Butler et al. manually dissected imaginal discs and were able to identify transcripts that were enriched in specific compartments of the wing discs [9]. However, these studies yield little information about the precise spatial patterns of gene expression.

Recent studies of precise spatial patterns of gene expression for large numbers of genes in developing

Drosophila embryos through in situ hybridization [22, 7] require annotation, and suffer from the fact that the annotation of spatial expression pattern requires manual curation. Kumar et al. [15] applied machine vision techniques to low-resolution images of in situ stained embryos and developed an algorithm for searching a database of patterns of gene expression in the embryos. Peng and Myers [21] have performed automated embryo registration and stain classification by using Gaussian mixture models. Yet, most of the previous work makes some parametric assumptions on the shape of the tissue, and the registration techniques used are very simplistic (such as aligning the major and minor axes for embryo images [21]). Shape learning and alignment is a well studied problem in the computer vision community. In our work, we chose *congealing* [17] since we are interested in aligning the binary shape masks that we obtain after segmentation and *congealing* is readily applicable. For further discussion on relevant shape alignment algorithms, reader is referred to [10, 23, 13, 17].

3.2 Proposed Approach

Figure 3.2 illustrates the data flow in our approach. In this report, we will only discuss the computational aspects related to the segmentation, alignment and shape learning.

Let us first provide a simple overview of the procedure. Starting with the a set of imaginal disc images of a given tissue class, we find the structure of interest in the given image via segmentation (Section 3). For imaginal discs, this means recognizing and extracting the portion of the image that corresponds to the imaginal disc and separating this from the cluttered image background. We use a combination of manual and automatic segmentation schemes to segment a set of these imaginal disc images and input these into the unsupervised shape learning algorithm called ‘*Congealing*’ [17]. This data-driven algorithm (*congealing*) learns the canonical shape model of a given ensemble of shapes and the set of transformations associated with each shape in the ensemble simultaneously by solving a constrained optimization problem (Section 4). *Congealing* operates by minimizing the summed component-wise (pixel-wise) entropies over a continuous set of transformations on the image ensemble and is robust against local minima in the optimization (Section 4.1). Once the underlying shape model is learned, it can be used subsequently as the canonical model. We also use this model as a shape template to improve our automatic segmentation algorithm (Section 3). The learned transformations are applied back to the original imaginal disc images to bring them into alignment in one step (Section 4.2). Once the images are aligned, we use a simple stain scoring algorithm (Section 5) to make quantitative analysis of the spatial expression patterns in these imaginal discs. These aligned stain scores are used as descriptors to make comparative analysis of spatial gene expression across multiple genes.

3.2.1 Image Model for *Drosophila* Imaginal Discs

Imaginal discs have a morphology similar to that of a flattened balloon. Imaginal discs do have substantial depth to them, but we image a single plane from the discs and consider an idealized 2-dimensional representation of a disc. A single focal plane is selected for each image to maximize visibility of any stain present in the disc and the choice of focal plane generally does not effect the

perceived boundaries of the disc.

Mathematically, we denote the set of input imaginal disc images of a given class as $\Phi \doteq \{I^i\}_{i=1}^N$ where N is the cardinality of the set. Each image $I^i(\cdot)$ can be represented as a map from the image \mathbb{R}^2 to the color space $C \subset \mathbb{R}^3$ with a small support $\Omega \subset \mathbb{R}^2$:

$$I^i(\mathbf{x}) : \mathbf{x} \in \Omega \mapsto \mathbf{c} = I^i(\mathbf{x}) \in C; \quad (3.1)$$

where $\mathbf{x} = [x, y, 1]^T$ (in homogeneous coordinates), $\mathbf{c} = [r, g, b]^T \in \mathbb{R}^3$ is a vector in the color space. In general, the domain Ω is a square or a rectangular window. This representation will be used throughout the rest of the chapter whenever the mathematical details of individual steps are explained.

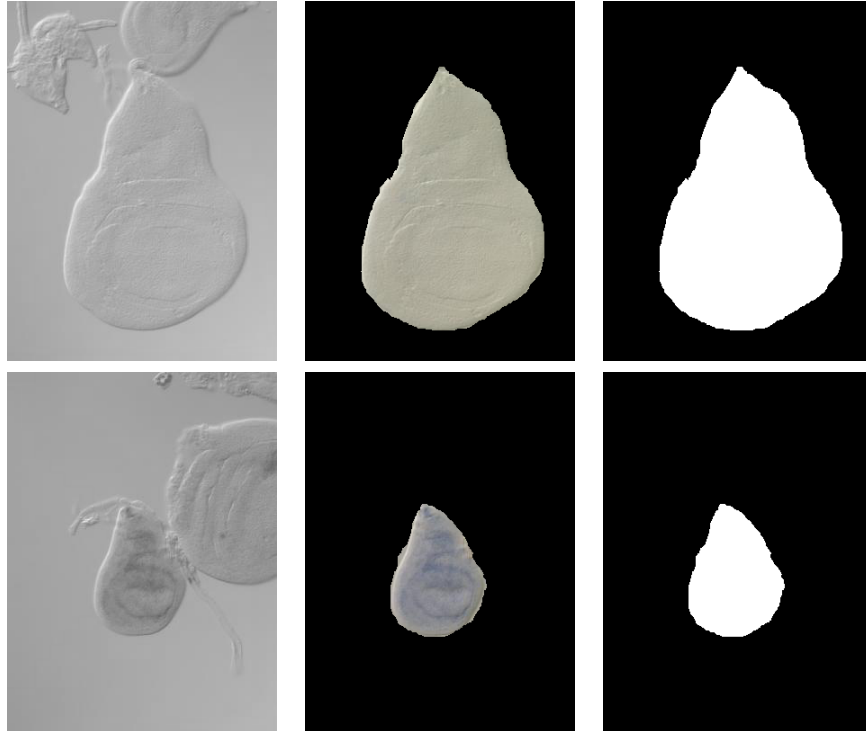


Figure 3.3: Example segmentation results using the combined segmentation procedure.

Example segmentation results for wing discs (*first row*) and haltere discs (*second row*): **Left Column:** Original image $I^i(\mathbf{x})$. **Middle Column:** Segmented tissue structure of interest $I_f^i(\mathbf{x})$. **Right Column:** Extracted binary shape $I_s^i(\mathbf{x})$.

3.3 Extracting Tissue Shapes via Segmentation

Two salient features of our image dataset make the segmentation task relatively simple. First, Nomarski images of the discs [22] yield substantial highlights and lowlights at the periphery of the discs. The background of the images is generally uniform and one can use the significant

contrast generated at the edge of the discs to identify border regions. Second, compared to the background, the pixel intensities of the imaginal disc tissues have much more variability than the background, even over a small window. While the mean pixel intensities of large regions of disc and background are very similar, the disc has a broader range of pixel intensities and, more importantly, the local derivative values are much higher for disc than background. This enables the use of either the magnitude of the gradient and/or the variance of a window around a pixel as a feature for distinguishing disc from non-disc tissue. This bimodal distribution lends itself to fitting a mixture of two Gaussian random variables followed by labeling of each pixel as disc or background. Furthermore, the second derivative of the image (the Laplacian) is also quite high at the edge of the disc and serves as a useful feature for identifying discs. Using these insights, we implemented a simple filter-and-threshold module for segmentation. It computes the *local variance* of the image in a small support region, estimates the bimodal distribution of variance in the filtered image and thresholds it appropriately to separate the disc region from the background. This process creates a binary shape mask that we use in following sections to learn the canonical shape model of the disc tissue.

Mathematically, this can be written as follows:

$$var(I^i(\mathbf{x})) = \int \int \mathbb{E}\{I^i(\mathbf{x})^2\} - (\mathbb{E}\{I^i(\mathbf{x})\})^2 dx dy \quad (3.2)$$

where $\mathbf{x} = [x, y, 1]^T$ (in homogeneous coordinates), and $\mathbb{E}(\cdot)$ is the expectation over the small square window centered at \mathbf{x} such that $-\alpha \leq x \leq +\alpha$, $-\alpha \leq y \leq +\alpha$, $\alpha \in \mathbb{R}$. The extracted shape image I_s^i is then calculated as:

$$I_s^i(\mathbf{x}) = \begin{cases} 1, & \text{if } var(I^i(\mathbf{x})) \geq \delta \\ 0, & \text{otherwise} \end{cases} \quad (3.3)$$

where $I_s^i(\mathbf{x})$ is a binary image of the extracted shape, $\mathbf{x} \in \Omega$ and δ is a threshold value where $\delta \in \mathbb{R}$. $I_s^i(\mathbf{x})$ is a map from the Ω to the set $B = \{0, 1\}$. The segmented structure of interest (or the foreground) $I_f^i(\mathbf{x})$ can be computed by point-wise multiplication of $I^i(\mathbf{x})$ and $I_s^i(\mathbf{x})$:

$$I_f^i(\mathbf{x}) = I^i(\mathbf{x}) \cdot I_s^i(\mathbf{x}). \quad (3.4)$$

where $I_f^i(\mathbf{x})$ is a map from Ω to C . In other words,

$$I_f^i(\mathbf{x}) = \begin{cases} I^i(\mathbf{x}), & \text{if } var(I^i(\mathbf{x})) \geq \delta. \\ 0, & \text{otherwise.} \end{cases} \quad (3.5)$$

Sometimes this simple filter-and-threshold process results in unsatisfactory performance, for two reasons. First, the presence of non-disc tissue (such as the trachea which is attached to the wing disc), may interfere with the segmentation of the disc from the background, thereby corrupting the extracted shape with the presence of this additional biological material. Second, some regions inside the disc appear more homogeneous than others and sometimes get misclassified as the background. This is especially true when the imaginal discs are heavily stained. Both these problems can be addressed by performing a template matching operation to identify the disc. Given the rough shape template of the disc, the template matching operation is quite simple but the problem is

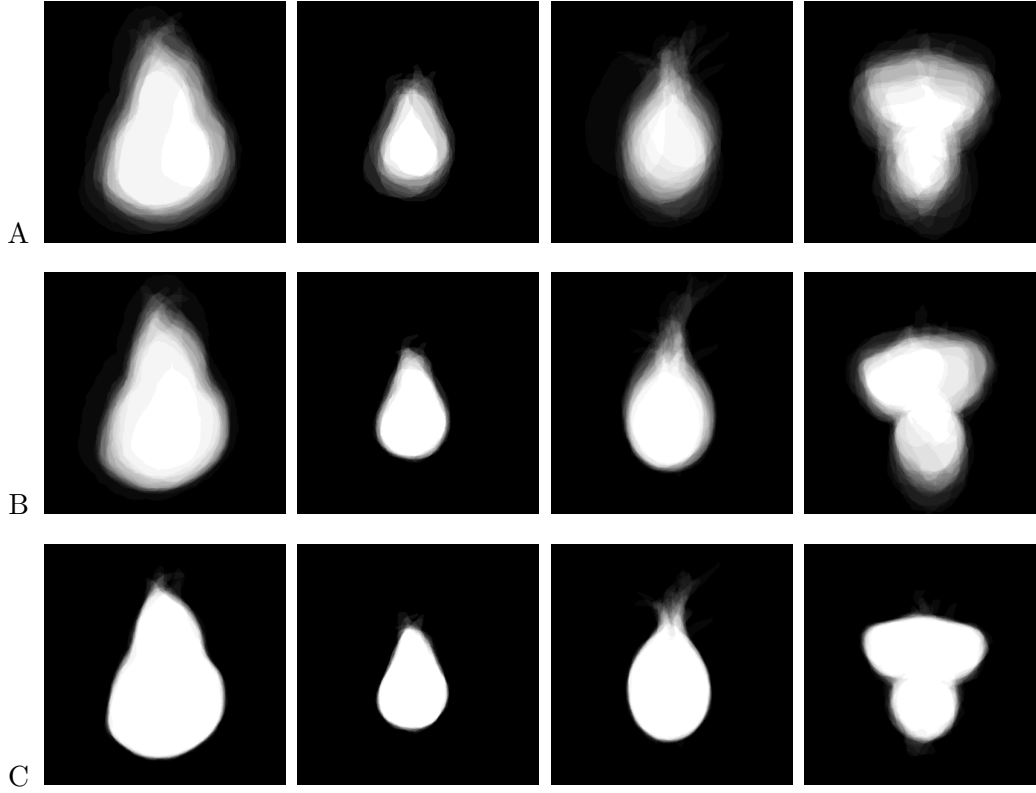


Figure 3.4: Mean shape images from the learning stage.

Mean images from optimization process during *Congeling* for wing discs (*first column*), haltere discs (*second column*) leg discs (*third column*), eye discs (*fourth column*): **A**: Mean image of Φ_s before congealing. **B**: Mean image of Φ_s after congealing to convergence with only 3 parameters (t_x, t_y and θ). **C**: Mean image of Φ_s after congealing to convergence with 7 parameters (Equation 3.10).

that there is no such clean shape template to begin with. We address this issue as follows: Using manual segmentation on a set of disc images, we obtain relatively clean shapes of the tissue for each class. These relatively clean structures are then fed to the nonparametric shape learning algorithm to form a good canonical shape template. This learned shape template was used in conjunction with the simple filter-and-threshold algorithm to obtain better segmentation results automatically in cluttered images. The manually segmented shapes were also used as truth data for comparing the performance of our implementations of segmentation algorithms. Our current implementation gives satisfactory segmentation results. We show some sample segmentation results from our segmentation procedure in Figure 3.3 for wing and haltere discs.

3.4 Joint Nonparametric Shape Learning and Alignment

Once the shapes of the relevant disc tissues, $I_s^i(\mathbf{x})$ (Equation 3.3), are extracted in binary image format, we use this set of binary shapes to learn the canonical underlying shape model of the given class of disc tissues using a nonparametric learning algorithm called ‘*Congealing*’ [17]. We denote the set of binary shape images as $\Phi_s \doteq \{I_s^i\}_{i=1}^N$ where N is the cardinality of the set. For a thorough discussion of this algorithm, the reader is referred to Miller et al. [17, 19].

In this section, we will explain the mathematical formulation of *Congealing* in the light of our image model for the imaginal discs.

Let us denote the latent binary shape of the given class of disc tissues as I_l . We model each shape image in Φ_s as I_l transformed through a geometric transformation. Given a class, the latent shape and the transformation are conditionally independent [19]. We assume that the transformations are affine and model the affine parameters as i.i.d. random variables. We shall assume that the transformation is a one-to-one and invertible mapping between I_l and I_s^i . We make the further assumption that the probability distribution of pixel values at each pixel location are *i.i.d.* Thus, for any given pixel location $\mathbf{x} = \mathbf{x}^*$,

$$I_s^i(\mathbf{x}^*) = I_l(g_i^{-1}(\mathbf{x}^*)). \quad (3.6)$$

In our implementation, we parameterize the set of transformations g_i using the following component transformations: x -translation (t_x), y -translation (t_y), rotation (θ), x -log-scale (s_x), y -log-scale (s_y), x -shear (h_x), and y -shear (h_y). Clearly, this is an *over-complete* parameterization, but following the efficiency arguments presented by Miller [19] it allows faster convergence during the optimization routine. We experimented with different choices of parameterization, and we will show results based on the parameterization as shown in Equation 3.10.

Fixing the order of composition to ensure unique mapping (since the matrix multiplication is not commutative), this can be written as:

$$g = F(t_x, t_y, \theta, s_x, s_y, h_x, h_y) \quad (3.7)$$

$$g_i = F(\{v_j\}^i) \quad (3.8)$$

$$\{v_j\}^i = (t_x^i, t_y^i, \theta^i, s_x^i, s_y^i, h_x^i, h_y^i) \quad (3.9)$$

where $1 \leq i \leq N$, (N is the cardinality of the set Φ_s), $1 \leq j \leq K$, (K is the number of parameters chosen), and $v \in \mathbb{R}^{N \times K}$.

Writing g out explicitly, we get:

$$g = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} e^{s_x} & 0 & 0 \\ 0 & e^{s_y} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & h_x & 0 \\ h_y & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.10)$$

3.4.1 Shape Learning

The goal is to find the transformation g_i that converts a given I_s^i into the most likely form of I_l . Formulating this as the maximum likelihood estimation problem, and by searching through the affine parameter space as defined above, we want to find g_i that maximizes $P(g_i^{-1}|I_s^i)$. We define $\hat{\Theta}$ as:

$$\hat{\Theta} = \arg \max_{g_i} P(g_i^{-1}|I_s^i). \quad (3.11)$$

Using Bayes' rule and ignoring the constant term in the denominator:

$$\hat{\Theta} = \arg \max_{g_i} P(I_s^i|g_i^{-1})P(g_i^{-1}). \quad (3.12)$$

Assuming uniform prior over the transformation parameters (Equation 3.8), we can write Equation 3.12 as:

$$\hat{\Theta} = \arg \max_{g_i} P(I_s^i|g_i^{-1}). \quad (3.13)$$

Notice the fact that probability of the I_s^i is the same as probability of I_l associated with the corresponding I_s^i and g_i^{-1} . We can write this as:

$$\hat{\Theta} = \arg \max_{g_i} P(I_l^i(I_s^i, g_i^{-1})). \quad (3.14)$$

Using our *i.i.d.* assumptions and Equations 3.6, 3.8 and 3.9,

$$\hat{\Theta} = \arg \max_v \prod_{\mathbf{x} \in \Omega} \prod_{v \in \mathbb{R}^{N \times K}} p_{\mathbf{x}}(I_s(g_i(\mathbf{x}))). \quad (3.15)$$

Taking log-probabilities, Equation 3.15 becomes:

$$\hat{\Theta} = \arg \max_v \sum_{\mathbf{x} \in \Omega} \sum_{v \in \mathbb{R}^{N \times K}} \log p_{\mathbf{x}}(I_s(g_i(\mathbf{x}))). \quad (3.16)$$

I_s^i is a binary image so the pixel stack at \mathbf{x} would consist of 1's and 0's. $p_{\mathbf{x}}(I_s(g_i(\mathbf{x})))$ is the empirical probability of 1's in the pixel stack at \mathbf{x} . Since we model the transformations as the random variables causing $I_s^i(\mathbf{x})$ to vary from $I_l(\mathbf{x})$, we can see that these two will be the same when the randomness due to g_i is removed. Substituting Shannon entropy to indicate the sample expectation of the function $-\log p_{\mathbf{x}}$ (Equation 3.19), we get:

$$\arg \max_v \sum_{\mathbf{x} \in \Omega} \sum_{v \in \mathbb{R}^{N \times K}} \log p_{\mathbf{x}}(I_s(g_i(\mathbf{x}))) \approx \arg \min_v \sum_{\mathbf{x} \in \Omega} H(g(\mathbf{x})). \quad (3.17)$$

By the law of large numbers, this approximation becomes equality when N is very large. Here, $H(\mathbf{x})$ is the Shannon entropy of the binary pixel stack at \mathbf{x} , defined by

$$H(g(\mathbf{x})) = -p_{\mathbf{x}} \log p_{\mathbf{x}} - (1 - p_{\mathbf{x}}) \log(1 - p_{\mathbf{x}}) \quad (3.18)$$

and

$$H(\mathbf{x}) = \mathbb{E}_p\{-\log p_{\mathbf{x}}\} \quad (3.19)$$

where $p_{\mathbf{x}} \doteq p_{\mathbf{x}}(I_s(g(\mathbf{x})))$. Mathematically, this ML estimation can be formulated as solving an optimization problem. Noting the fact that the priors aren't actually uniform on different

transformations, we need to write a compensated objective function for our optimization Ψ defined as

$$\Psi \doteq \sum_{\mathbf{x} \in \Omega} H(g(\mathbf{x})) + \sum_{i=1}^N |v^i| \quad (3.20)$$

where v^i are the vectors of transformation parameters (Equation 3.9), and $|\cdot|$ is some norm (or penalty term or regularization term) on these vectors to keep the shape images from shrinking to zero or undergoing some other extreme transformations. Ψ is called the penalized pixel-wise entropy [19].

The learning algorithm proceeds as follows:

1. Maintain a transform parameter vector v^i (Equation 3.9) for each shape image I_s^i . Each parameter vector will specify a transformation matrix $\hat{g}_i = F(v^i)$ according to Equation 3.10. Initialize all v^i to zero vectors. This has the effect of initializing all of the transformation matrices \hat{g}_i to the identity matrix.
2. Compute the penalized pixel-wise entropy Ψ for the current set of images from Equation 3.20.
3. Repeat until convergence:

For each shape image I_s^i ,

- (a) Calculate the numerical gradient $\nabla_{v^i} \Psi$ of Equation 3.20 with respect to the transformation parameters v_j^i 's for the current image ($1 \leq j \leq K$).
- (b) Update v^i as: $v^i = v^i + \gamma \nabla_{v^i} \Psi$. (where the scaling factor $\gamma \in \mathbb{R}$).
- (c) Update γ (according to some reasonable update rule such as the Armijo rule [8]).

At convergence of this optimization procedure, the set of shapes $\Phi_s = \{I_s^i\}_{i=1}^N$ are all aligned, and the associated transformations $\{g_i\}_{i=1}^N$ are computed. To visualize the entropy of the transformed image set for a class at each step of the optimization, one can construct an image (Figure 3.4) in which each pixel is the mean of its corresponding pixel stack.

3.4.2 Joint Alignment

We apply the $\{v_j\}^i$ learned from the congealing process to the extracted structures I_f^i to bring all the images into alignment in one step.

$$I_a^i(\mathbf{x}) = I_f^i(g_i(\mathbf{x})). \quad (3.21)$$

where $1 \leq i \leq N$. We show our results in Figures 3.5, 3.6, 3.7, 3.8.

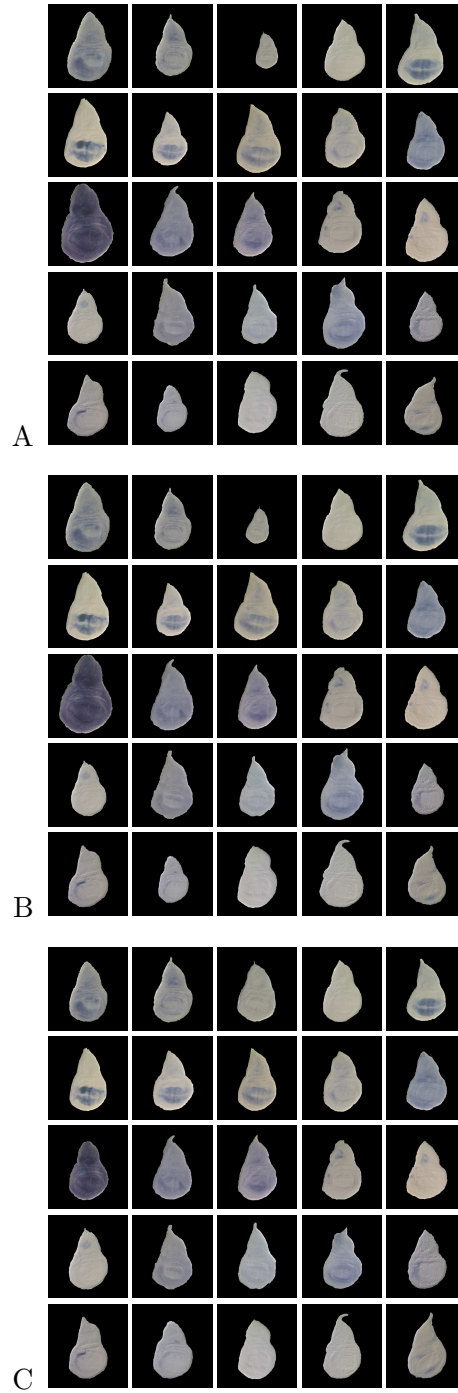


Figure 3.5: Alignment results for wing discs.

A: Segmented wing discs I_f^i before congealing. **B:** Segmented wing discs I_a^i after applying the transformations learned by congealing with only 3 parameters (t_x, t_y and θ). **C:** Segmented wing discs I_a^i after applying the transformations learned by congealing with 7 parameters (Equation 3.10). Note that the third image was taken with a 10x objective, while the rest were taken with a 20x objective. The reduced size of the disc is due to the way we captured the image, rather than due to biological variability. However, the algorithm was able to properly align the image as shown in **C**.

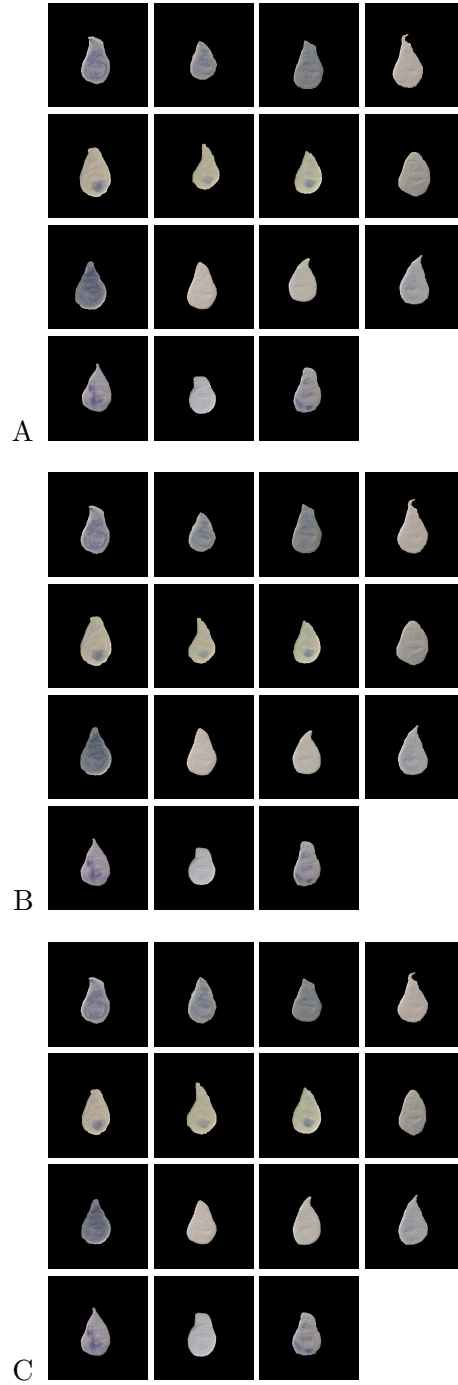


Figure 3.6: Alignment results for haltere discs.

A: Segmented haltere discs I_f^i before congealing. **B:** Segmented haltere discs I_a^i after applying the transformations learned by congealing with only 3 parameters (t_x, t_y and θ). **C:** Segmented haltere discs I_a^i after applying the transformations learned by congealing with 7 parameters (Equation 3.10).

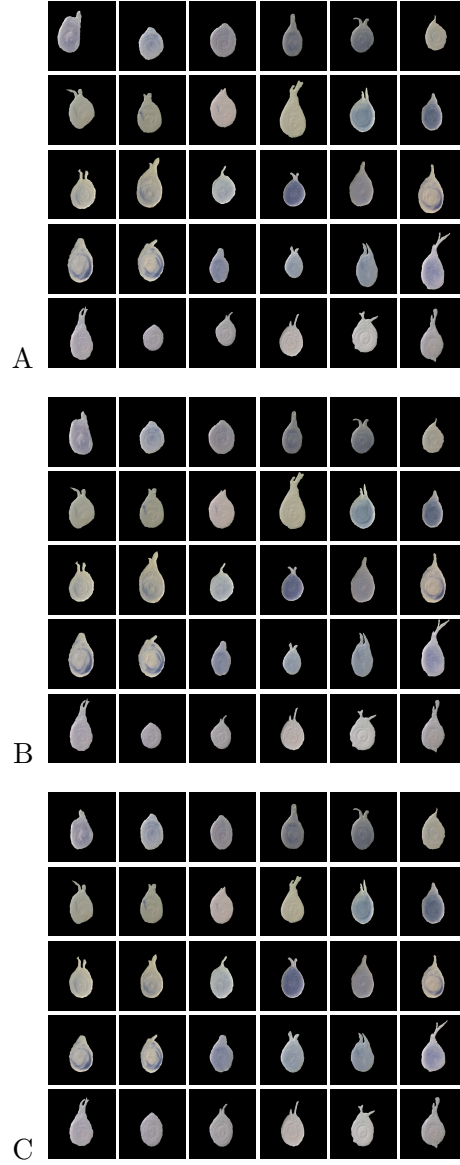


Figure 3.7: Alignment results for leg discs.

A: Segmented leg discs I_f^i before congealing. **B:** Segmented leg discs I_a^i after applying the transformations learned by congealing with only 3 parameters (t_x, t_y and θ). **C:** Segmented leg discs I_a^i after applying the transformations learned by congealing with 7 parameters (Equation 3.10).

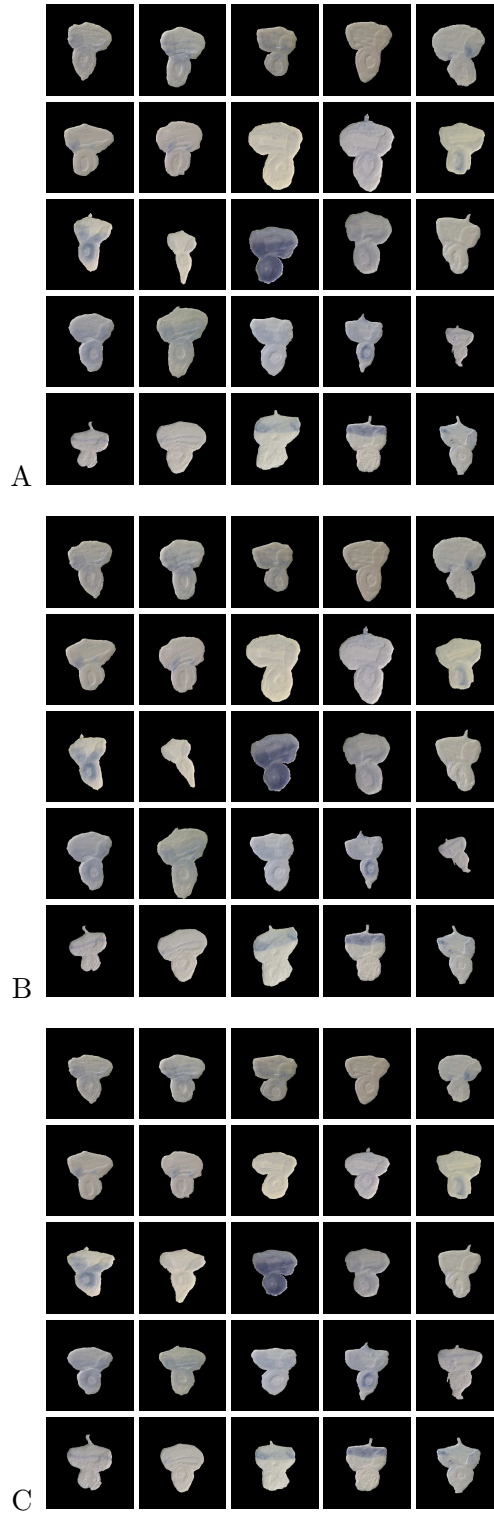


Figure 3.8: Alignment results for eye discs.

A: Segmented eye discs I_f^i before congealing. **B:** Segmented eye discs I_a^i after applying the transformations learned by congealing with only 3 parameters. **C:** Segmented eye discs I_a^i after applying the transformations learned by congealing with 7 parameters (Equation 3.10).

3.5 Stain Scoring

Mass-isolated imaginal discs were placed in 96-well plates and individual wells were stained with digoxigenin-labeled RNA or DNA complementary to genes of interest. Images were acquired using a light microscope equipped with Nomarski optics as described by Tomancak et al. [22].

The local presence of stain results in the appearance of blue in the image; darker blue suggests a greater local concentration of the gene of interest. However, there is substantial probe-to-probe variability and these intensities should not be relied on as an accurate quantitative measure of gene concentration. Nevertheless, the different intensity values can be used to suggest where local gene concentration is high.

We developed a simple semi-quantitative metric that ranges from 0 to 5 where 0 indicates no expression of the gene of interest and 5 indicates high expression; the presence of blue stain causes a decrease in the intensities of the red and green channels in an RGB image. In unstained discs the Nomarski optics yield an imaginal disc image that is generally varying shades of gray, where the local features and folds of the tissue result in lighter and darker intensities. The intensities of the red, blue and green channels are generally in the same range for a given pixel. We measured the intensity of the blue channel minus the average the red and green channels to determine the level of staining. A small baseline value was subtracted from this number to reduce local noise due to the variability of the intensities of the channels of unstained images and the resulting value was then thresholded into six values with the lowest value suggesting no stain and, therefore, no or minimal gene of interest present, and the highest value suggesting strong expression of the gene of interest.

Examples of segmented, stain-scored wing discs, both unadjusted and congealed, can be seen in Figure 3.9. Notice that the overall shape and size of the discs are more consistent in the congealed images and that a pixel-wise comparison of stain intensity of biologically similar patterns would appear more similar when comparing the congealed, stain-scored images than the unaligned stain-scored images.

3.6 Summary and Future Work

The proposed overall methodology shown in Figure 3.2 operates without making any assumptions about the underlying structure of a given tissue class. It is unsupervised and is highly amenable to large scale spatial expression analysis. It augments any model-based registration methods one may choose to apply by supplying the nonparametrically learned canonical structure model from the given ensemble of images for a given tissue class. We successfully implemented and demonstrated the applicability of this methodology using *Drosophila* imaginal discs. We presented a mathematical and probabilistic analysis of the nonparametric shape learning algorithm used. Using the salient features of our images, we suggested a simple filter-and-threshold algorithm for segmentation which performs well once the learned shape template is supplied. However, this simple segmentation module may not work well for different types of imaging processes. To make our approach more general, we are currently investigating the possibility of incorporating more refined segmentation algorithms into our approach. The discussion in this chapter is focused on the geometric transformations that can be approximated by an affine model in a two-dimensional plane

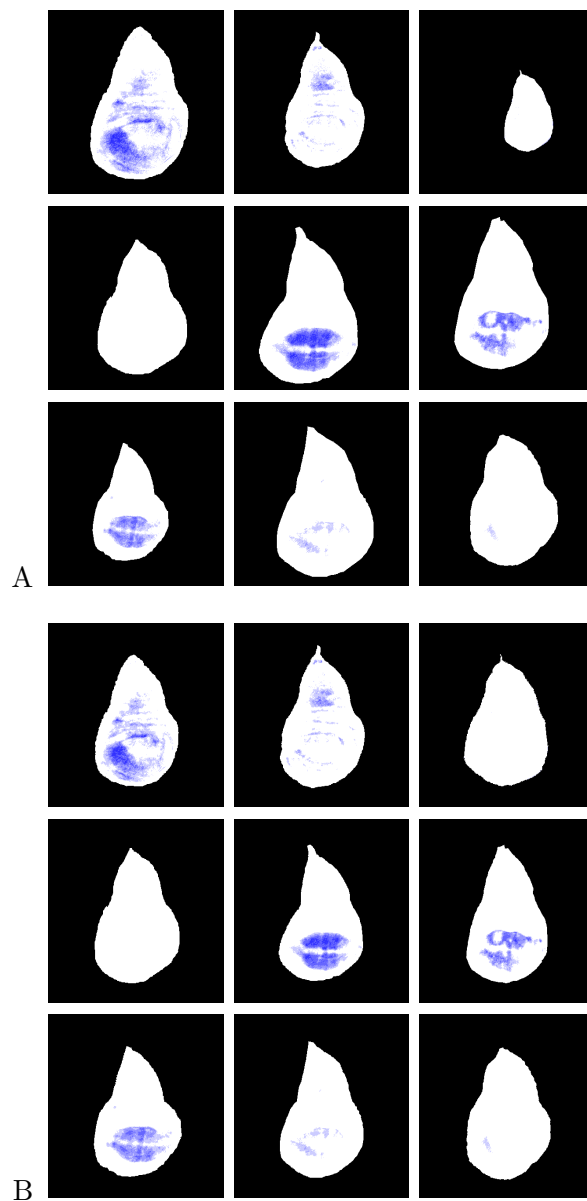


Figure 3.9: Stain patterns in *Drosophila* imaginal disc images: unaligned vs. aligned. **A:** Unaligned stain-scored wing disc images. **B:** Aligned stain-scored wing disc images after congealing. Black pixels have been segmented as background, white indicates no stain and shades of blue indicate stained pixels. Blue intensity values were calculated from the semi-quantitative stain scoring algorithm with the lightest blue representing value 1 and the darkest blue representing value 5. It can be noted that pixel-wise stain count is much more meaningful in these images. (This image is better viewed in color).

since imaginal discs can be represented using a two-dimensional representation. In future work we plan to extend this approach to three-dimensional datasets such as image stacks from confocal microscopy studies of in situ stained tissues (such as *Drosophila* embryos). Furthermore, we plan to perform detailed comparative analyses of spatial patterns of gene expression in aligned imaginal discs using pixel-wise comparisons based on the approaches described in this work.

Bibliography

- [1] Brain web project. <http://www.bic.mni.mcgill.ca/brainweb/>.
- [2] Parvez Ahammad, Cyrus Harmon, Ann Hammonds, Shankar Sastry, and Gerald Rubin. Joint nonparametric alignment for analyzing spatial gene expression patterns of drosophila imaginal discs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [3] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, Jr., L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt. Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling. *Nature*, 403:503–511, 2000.
- [4] M. N. Arbeitman, E. E. M. Furlong, F. Imam, E. Johnson, B. H. Null, B. S. Baker, M. A. Krasnow, M. P. Scott, R. W. Davis, and K. P. White. *Science*, 297:2270–2275, 2002.
- [5] Michael Bate and Alfonso Martinez Arias. *The development of Drosophila melanogaster*, volume II. 1993.
- [6] J. Beirlant, E. Dudewicz, L. Györfi, and E. van der Meulen. Nonparametric entropy estimation: An overview. *International Journal of Mathematical and Statistical Sciences*, 6:17–39, 1997.
- [7] BP. Berman, Y. Nibu, BD. Pfeiffer, P. Tomancak, SE. Celniker, M. Levine, GM. Rubin, and MB. Eisen. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the drosophila genome. *Proc Natl Acad Sci U S A*, 99(2):757–62, 2002.
- [8] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.
- [9] Miranda J. Butler, Thomas L. Jacobsen, Donna M. Cain, Michael G. Jarman, Michael Hubank, J. Robert S. Whittle, Roger Phillips, and Amanda Simcox. Discovery of genes with highly restricted expression patterns in the Drosophila wing disc using DNA oligonucleotide microarrays. *Development*, 130:659–670.
- [10] G. Charpiat, O. Faugeras, and R. Keriven. Shape metrics, warping and statistics. In *Proceedings of the International Conference on Image Processing*, 2003.

- [11] D.L. Collins, A.P. Zijdenbos, J.G. Kollokian, N.J. Sled, C.J. Kabani, C.J. Holmes, and A.C. Evans. Design and construction of a realistic digital brain phantom. *IEEE Transactions on Medical Imaging*, 17:463–468, 1998.
- [12] Ayres Fan. A variational approach to MR bias correction. *M.S. Thesis*, Massachusetts Institute of Technology, 2003.
- [13] B. Frey and N. Jojic. Learning mixture models of images and inferring spatial transformations using the em algorithm. In *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition*, 1999.
- [14] A. Klebes, B. Biehls, F. Cifuentes, and TB. Kornberg. Expression profiling of drosophila imaginal discs. *Genome Biol*, 3(8):RESEARCH0038, 2002.
- [15] Sudhir Kumar, Karthik Jayaraman, Sethuraman Panchanathan, Rajalakshmi Gurunathan, Ana Marti-Subirana, and Stuart J. Newfeld. BEST: a novel computational approach for comparing gene expression patterns from early stages of drosophila melanogaster development. *Genetics*, 162:2037–2047, 2002.
- [16] Erik G. Learned-Miller and Parvez Ahammad. Joint MRI bias removal using entropy minimization across images. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems 17*. MIT Press, Cambridge, MA, 2005.
- [17] Erik G. Learned-Miller, Nicholas Matsakis, and Paul A. Viola. Learning from one example through shared densities on transforms. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2000.
- [18] R.J. Lipshutz, SP. Fodor, TR. Gingeras, and DJ. Lockhart. High density synthetic oligonucleotide arrays. *Nat Genet*, 21(1 Suppl):20–4, 1999.
- [19] Erik G. Miller. Learning from one example in machine vision by sharing probability densities. *PhD Dissertation*, Massachusetts Institute of Technology, 2002.
- [20] Dwight Nishimura. *Principles of Magnetic Resonance Imaging*. Stanford University, Palo Alto, CA, 1996.
- [21] Hanchuan Peng and Eugene W. Myers. Comparing in situ mrna expression patterns of drosophila embryos. In *Proceedings of the Eighth Annual International Conference on Research in Computational Molecular Biology*, 2004.
- [22] Pavel Tomancak, Amy Beaton, Richard Weiszmam, Elaine Kwan, ShengQiang Shu, Suzanna E Lewis, Stephen Richards, Michael Ashburner, Volker Hartenstein, and Gerald M Rubin. Systematic determination of patterns of gene expression during drosophila embryogenesis. *Genome Biology*, 3(12):1–14.
- [23] A. Tsai et al. A shape-based approach to curve evolution for segmentation of medical imagery. *IEEE Transactions on Medical Imaging*, 22, No. 2:137–154, February 2003.
- [24] O. Vasicek. A test for normality based on sample entropy. *Journal of the Royal Statistical Society Series B*, 31:632–636, 1976.

- [25] Paul A. Viola. Alignment by maximization of mutual information. *PhD Dissertation*, Massachusetts Institute of Technology, 1995.
- [26] W. M. Wells, W. E. L. Grimson, R. Kikinis, and F. Jolesz. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 15:429–442, 1996.